

Topic 12. The Split-Plot and its Relatives [S&T Ch 16]

12. 1. Definition

The split-plot design results from a specialized randomization scheme for a factorial experiment.

The basic split-plot design involves assigning the treatments of one factor to **main plots** arranged in a CRD, RCBD or a Latin-Square design and then assigning the second factor to **subplots** within each main plot.

Note that randomization is a **two-stage** one. First, levels of factor A are randomized over the main plots and then levels of factor B are randomized over the subplots. Each main plot may be considered as a block as far as factor B is concerned but only as an **incomplete block** as far as the full set of treatments is concerned because not every subplot has the same chance of getting every treatment combination.

This restriction in randomization results in the presence of **two error terms**, one for main plots and one for subplots. Ordinarily the error term for the main plots is larger than it would be in a complete design since the main plots are larger and further apart, while the subplot error is smaller than it would be in a complete design. Since the interactions are compared using the smaller subplot error, the precision in estimating interactions is usually increased.

A classical example of a split plot is an irrigation experiment where irrigation levels are applied to large areas, and factors like varieties and fertilizers are assigned to smaller areas within a particular irrigation treatment. The proper analysis of a split-plot design recognizes that treatments applied to main plots are subject to larger experimental error than those applied to subplots; hence, different mean squares are used as denominators for the corresponding F ratios. This concept is discussed in terms of expected mean squares in this topic.

In summary, the factors that require smaller amount of experimental material, that are of major importance, that are expected to exhibit smaller differences, or for which greater precision is desired are assigned to the subplots.

12. 2. Uses of Split-plot designs

- 1) Split-plot designs, and a variation, the split-block, are frequently used for factorial experiments in which the nature of the experimental material or the operations involved make it difficult to handle all factor combinations in the same manner. It may be used when the treatments associated with the levels of one or more of the factors require larger amounts of experimental material in an experimental unit than do treatments for other factors.
- 2) These designs are also used when the investigator wishes to increase precision in estimating certain effects and is willing to sacrifice precision in estimating certain others. The design usually sacrifices precision in estimating the average effects of the treatments assigned to main plots. It often improves the precision for comparing the average effects of treatments assigned to subplots and, when interactions exist, for

comparing the effects of subplot treatments for a given main plot treatment. This arises from the fact that experimental error for main plots is usually larger than the experimental error used to compare subplot treatments. Usually, the error term for subplot treatments is smaller than would be obtained if all treatment combinations were arranged in a randomized complete block design.

- 3) The design may be used when an additional factor is to be incorporated in an experiment to increase its scope. For example, suppose that the major purpose of an experiment is to compare the effects of several seed protectants. To increase the scope of the experiment several varieties are used as main plots and the seed protectants are used as subplots.

12. 3. Split-Plot design

Suppose factor A is to be the main plot factor and is applied at 3 levels, while factor B is to be the subplot factor and is applied at 2 levels. There are 4 reps per main plot. We will see how this is arranged in 3 designs, factorial (no split) arranged as a CRD, split-plot with completely randomized main plots, and split-plot with randomized complete block main plots.

12. 3. 1. Factorial (no split) with completely randomized treatment combinations.

There are 6 combinations times 4 reps = 24 plots. The six treatment combinations are applied randomly across the plots.

a1b1	a2b2	a2b1	a1b2	a3b2	a1b1	a2b2	a2b1	a1b2	a3b2	a1b1	a3b2
a2b2	a3b1	a1b2	a3b1	a1b2	a3b2	a2b1	a1b1	a2b2	a3b1	a2b1	a3b1

12. 3. 2. Split-plot with completely randomized main plots. Randomization is divided into 2 stages;

Stage 1: Randomize the levels of factor A over the main plots.

a2	a3	a2	a1	a2	a3	a2	a3	a1	a3	a1	a1
a2	a3	a2	a1	a2	a3	a2	a3	a1	a3	a1	a1

Stage 2: Randomize the levels of B over the subplots, 2 per subplot.

a2b2	a3b2	a2b1	a1b1	a2b1	a3b2	a2b1	a3b2	a1b1	a3b1	a1b1	a1b2
a2b1	a3b1	a2b2	a1b2	a2b2	a3b1	a2b2	a3b1	a1b2	a3b2	a1b2	a1b1

12. 3. 3. Split-plot with randomized complete block main plots.

Stage 1: Randomize the levels of factor A over the main blocks.

a2	a1	a3	a1	a2	a3	a1	a3	a2	a3	a2	a1
a2	a1	a3	a1	a2	a3	a1	a3	a2	a3	a2	a1

Stage 2: Randomize the levels of B over the subplots, 2 per subplot.

a2b1	a1b1	a3b2	a1b2	a2b1	a3b2	a1b2	a3b1	a2b2	a3b1	a2b1	a1b1
a2b2	a1b2	a3b1	a1b1	a2b2	a3b1	a1b1	a3b2	a2b1	a3b2	a2b2	a1b2

There is usually a positive correlation between b1 and b2 values, because of their proximity within each main plot. Since the variance of factor b is calculated from the sum of $(b1-b2)^2/df$, a positive correlation will result in a smaller variance among b levels than in a normal factorial design.

12. 4. The linear model for the split-plot

The linear model for the split-plot with completely randomized main plots is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ik} + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

where

$k = 1, \dots, r$ indexes the reps,

$i = 1, \dots, a$ indexes the main plot levels, and

$j = 1, \dots, b$ indexes the subplot levels.

The γ_{ik} represent the error associated with the main plots, and the ε_{ijk} the error associated with the subplots. The variance of the γ terms is σ^2_γ and the variance associated with the ε terms is σ^2_ε . Usually σ^2_γ is the larger of the two. The linear model for the split-plot with RCB main plots is

$$Y_{ijk} = \mu + \rho_k + \alpha_i + \beta_j + \gamma_{ik} + (\alpha\beta)_{ij} + \varepsilon_{ijk}.$$

The extra term ρ_k represents the block effect.

The sum of squares equations for the CRD are

$SS = SS(MP) + SS(SP)$, where

$SS(MP) = SSA + SS(MPE)$

$SS(SP) = SSB + SS(AxB) + SS(SPE)$

12. 5. Split-plot ANOVA

Total degrees of freedom in the experiment are one less than the total number of subplots. Total $df = rab - 1$, where r = number of replications, a = number of main plots and b = number of subplots per main plot. Main plot (factor A) SS has $df (MP) = a - 1$ and subplot (factor B) has $df (SP) = b - 1$.

Main plot error, MS (MPE), often called “error A”, is computationally equivalent to the *Main plots x Replication* interaction in a CRD and to the *Main plots x Block* interaction in a RCBD. This error term is the appropriate error term in testing for differences among main plots.

$$\text{Main plot error} = \text{Main plots} \times \text{Replication}$$

If the responses of the subplots within a main plot are averaged, the resulting design is a RCBD with one observation per block-treatment combination. Remember that in this case the appropriate error term is the *Block * Treatment* interaction (Topic 9.7.5.2). Therefore, it makes sense to use this error term in the split-plot to compare the main-plot effects.

Subplot error, MS (SPE), often called “error B”, is estimated from the interaction [*Subplot x Replication + Main plot x Subplot x Replication*]. This second error term is used to estimate the significance of the subplot effect and the subplot x main plot interaction effect. This is the residual error, or in other words the variation that is left after the variation of all the variables and interactions included in the model is extracted from the total variation.

$$\text{Subplot error} = \text{Subplot} \times \text{Replication} + \text{Main plot} \times \text{Subplot} \times \text{Replication}$$

The ANOVA table for the CRD split-plot is

Source	df	SS	MS	F
Main plots total	$ra - 1$	SS(MP)		
Factor A	$a - 1$	SSA	MSA	MSA/MS(MPE)
Main plot error	$a(r - 1)$	SS(MPE)	MS(MPE)	
Factor B	$b - 1$	SSB	MSB	MSB/MS(SPE)
A x B	$(a - 1)(b - 1)$	SS(AxB)	MS(AxB)	MS(AxB)/MS(SPE)
Subplot error	$a(r - 1)(b - 1)$	SS(SPE)	MS(SPE)	
Total (subplots)	$rab - 1$	SS		

The formulas for the split plot with main plots organized in RCBD or LS are similar and are given in Table 16.1 of ST&D (p.402). These different designs will not affect the last four rows of the previous table. The three upper lines are:

CRD		RCBD		Latin Square	
A	a-1	Blocks	r-1	Rows	a-1
Error A	a(r-1)	A	a-1	Columns	a-1
		Error A	(r-1)(a-1)	A	a-1
				Error A	(a-1)(a-2)
Total	ra-1	Total	ra-1	Total	ra-1
Factor B	b-1	Factor B	b-1	Factor B	b-1
A x B	(a-1)(b-1)	A x B	(a-1)(b-1)	A x B	(a-1)(b-1)
Error B	<u>a(r-1)(b-1)</u>	Error B	<u>a(r-1)(b-1)</u>	Error B	<u>a(r-1)(b-1)</u>
Total	rab-1	Total	rab-1	Total	rab-1

Error B (B*Block+A*B*Block) df = (b-1)*(r-1) + (b-1)*(r-1)*(a-1) = (b-1)*(r-1)*[1+(a-1)] = a*(b-1)*(r-1)

CRD

```
proc glm;
class rep A B;
model y= A rep*A B A*B;
test h=A e=rep*A;
```

RCBD

```
proc glm;
class bl A B;
model y=bl A bl*A B A*B;
test h=A e=bl*A;
```

LS

```
proc glm;
class row col A B;
model y=col row A col*row*A B A*B;
test h=A e= col*row*A;
```

Replicated Latin Squares (same rows and columns)

```
proc glm;
class sqr row col A B;
model y=sqr col row A sqr*col*row*A B A*B;
test h=A e= sqr*col*row*A;
```

12. 6. Example of a split-plot with randomized complete block main plots

To illustrate the technique, we will use an experiment of Thomson et al. (Phytopathology 71:605-608) to determine the effect of bacterial vascular necrosis on root yield of sugar beet at different in-row spacings. The two factors in the experiment were inoculation (inoculated versus not inoculated with *Erwinia carotovora*) and in-row spacing between plants (4, 6, 12, and 18 inches). The layout of this field experiment is shown in Figure 1. Note that in this experiment the bacterial inoculation was applied to large plots (main plot or whole plot) and the spacings were assigned to plots (subplots) within the main plots. The reasons for assigning inoculation to main plots were to avoid the contamination of healthy plants in non-inoculated plots during the process of inoculation, and because large differences in yield are expected between healthy and diseased plants.

The inoculation treatments were randomly assigned to the main plots within each of the six blocks. Then as far as the main plot treatments were concerned, this is a randomized complete block design. Where appropriate other designs such as Latin square can also be implied for the assignment of main plot treatments. The subplot treatments (spacing) are randomly assigned independently within each main plot. Fig. 1. Split plot layout of the sugar beet root rot study. Each block contains 2 main plots to which the inoculation treatments were assigned. Each main plot was split into 4 subplots to which the in-row spacing were assigned.

Block								
VI	4	12	18	6	6	12	4	18
	21.0	22.9	23.1	22.0	17.6	16.1	16.8	13.1
	No inoculation				Inoculation			
V	18	6	4	12	6	4	12	18
	12.9	19.8	17.2	16.8	21.2	17.9	22.3	22.0
	Inoculation				No inoculation			
IV	6	18	4	12	12	18	6	4
	21.1	21.4	18.4	22.8	16.1	14.7	16.3	16.8
	No inoculation				Inoculation			
III	18	12	4	6	18	6	12	4
	19.3	18.6	18.2	20.8	12.5	19.1	16.6	16.5
	No inoculation				Inoculation			
II	12	6	18	4	4	12	18	6
	14.9	17.0	12.1	16.4	17.9	21.1	20.1	19.6
	Inoculation				No inoculation			
I	4	12	18	6	18	12	6	4
	17.4	16.3	12.5	17.3	20.0	21.8	20.2	20.1
	Inoculation				No inoculation			

12. 6. 1. SAS Program

```

data split;
input A_inoc block B_space yield @@;
cards;
1 1 4 17.4 1 1 6 17.3 1 1 12 16.3 1 1 18 12.5 0 1 4 20.1 0 1 6 20.2 0 1 12 21.8 0 1 18 20.0
1 2 4 16.4 1 2 6 17.0 1 2 12 14.9 1 2 18 12.1 0 2 4 17.9 0 2 6 19.6 0 2 12 21.1 0 2 18 20.1
1 3 4 16.5 1 3 6 19.1 1 3 12 16.6 1 3 18 12.5 0 3 4 18.2 0 3 6 20.8 0 3 12 18.6 0 3 18 19.3
1 4 4 16.8 1 4 6 16.3 1 4 12 16.1 1 4 18 14.7 0 4 4 18.4 0 4 6 21.1 0 4 12 22.8 0 4 18 21.4
1 5 4 17.2 1 5 6 19.8 1 5 12 16.8 1 5 18 12.9 0 5 4 17.9 0 5 6 21.2 0 5 12 22.3 0 5 18 22.0
1 6 4 16.8 1 6 6 17.6 1 6 12 16.1 1 6 18 13.1 0 6 4 21.0 0 6 6 22.0 0 6 12 22.9 0 6 18 23.1
;
proc glm;
class block A_inoc B_space;
model yield=block A_inoc block*A_inoc B_space A_inoc*B_space;
test h=A_inoc e=block*A_inoc;
test h=block e=block*A_inoc;
means A_inoc;
means B_space;
means A_inoc*B_space;
run; quit;

```

The **test** statement specifies that the hypothesis concerning the main plots ($h = A_inoc$) is tested using the **block*A_inoc** interaction as the error term ($e = \text{blocks} * A_inoc$). This statement requests an F test whose numerator is the A_INOC mean square and whose denominator is the **block*A_inoc** mean square. The TEST statement is necessary in a split-plot analysis because all the default F tests use the residual mean square in the denominator, and that is not statistically valid when testing main-plot effects. It is not necessary to specify ERROR B because it is the residual.

12. 6. 2. SAS Output

```

Class      Levels  Values
BLOCK      6      1 2 3 4 5 6
A_INOC     2      0 1
B_SPACE    4      4 6 12 18
Number of observations in data set = 48

```

Dependent Variable: YIELD

Source	DF	SS	Mean Square	F Value	Pr > F
Model	17	388.5475	22.8557	29.17	0.0001
Error	30	23.5050	0.7835		
Corrected Total	47	412.0525			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	5	16.2500	3.2500	4.15	0.0055
A_INOC	1	256.6875	256.6875	327.62	0.0001
BLOCK*A_INOC	5	11.5350	2.3070	2.94	0.0280
B_SPACE	3	39.6375	13.2125	16.86	0.0001
A_INOC*B_SPACE	3	64.4375	21.4791	27.41	0.0001

Tests of Hypotheses using Type III MS for BLOCK*A_INOC as an error term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A_INOC	1	256.6875	256.6875	111.26	0.0001

Tests of Hypotheses using Type III MS for BLOCK*A_INOC as an error term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	5	16.2500	3.2500	1.41	0.3580

These F values indicate significant differences between *Inoculations*, between *Spacings* and a significant *Inoculation * Spacing* interaction. Note that when the appropriate error term is used the differences among blocks are not significant. Do not use the incorrect F value of 4.15 from the ANOVA Table, which declares differences among *Blocks*. ANOVA table of sugar beet root rot of Figure 1.

Source	df	SS	MS	F
Total (subplots)	47	412.06		
Block	5	16.26	3.25	1.41
Inoculation (A)	1	256.69	256.69	111.26
Error A (Block * A)	5	11.54	2.31	
Spacing (B)	3	39.64	13.21	16.86
Interaction (A x B)	3	64.44	21.48	27.41
Error B	30	23.50	0.78	

Interpretation: Note that the mean square for Error A is greater than the mean square for Error B. The coefficient of variation (CV) for main plots is 8.3% [$(\sqrt{2.31/18.26}) \times 100$] and for subplots is 4.8% [$(\sqrt{0.78/18.26}) \times 100$]. This is usual for split plot experiments. In the factorial experiment the interaction of the treatment factors is usually of primary importance, the smaller Error-B increases the chance of detecting true interactions. In this case, the interaction between inoculation and spacing is highly significant. This

indicates that the magnitude of the difference between inoculation treatments depends on spacing. For example, the difference for the 4 inches spacing between inoculated and non-inoculated plots was 2.1 tons/acre while 8-ton/acre difference was observed at 18-inch spacing. The significant interaction effect implies that the difference between levels of one factor depend on the other.

12. 6. 3. Appropriate error terms

To verify the appropriate tests the following RANDOM statement can be added to the previous program:

```
proc glm;
  class block A_inoc B_space;
  model yield=block A_inoc block*A_inoc B_space A_inoc*B_space;
  random block block * A_inoc;
```

The output of this statement is:

Source	Type III Expected Mean Square
BLOCK	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{BLOCK} * \text{A_INOC}) + 8 \text{Var}(\text{BLOCK})$
A_INOC	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{BLOCK} * \text{A_INOC}) + Q(\text{A_INOC}, \text{A_INOC} * \text{B_SPACE})$
BLOCK * A_INOC	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{BLOCK} * \text{A_INOC})$
B_SPACE	$\text{Var}(\text{Error}) + Q(\text{B_SPACE}, \text{A_INOC} * \text{B_SPACE})$
A_INOC * B_SPACE	$\text{Var}(\text{Error}) + Q(\text{A_INOC} * \text{B_SPACE})$

The expected means squares indicate that the appropriate error term for block and the main plot is the BLOCK * A_INOC interaction. The appropriate error term for B_SPACE and A_INOC * B_SPACE is the residual VAR(ERROR) term.

12. 6. 4. Mean comparisons

Mean separation in a split plot design is considerably more complicated than experiments which only involve a single error term. There are four different LSD's that can be calculated depending on the desired comparison:

If **interaction** between main plot * subplot is **not significant**

12. 6. 4. 1. Among main plot treatments.

12. 6 .4. 2. Among subplot treatments.

If **interaction** between main plot * subplot is **significant**

12. 6 .4. 3. Among subplot treatments within a main plot treatment.

12. 6 .4. 4. Among main plot treatments within a subplot.

12. 6 .4. 5. Among subplot treatments for different main plot treatments.

12. 6. 4. 1. Main plot comparisons in the absence of interaction

If no significant interactions are detected between main plot effect and the subplot effect, it is valid to compare each factor across all levels of the other factor. For didactic purposes we will use the previous example even though there was a significant interaction.

A valid comparison between means of the main plots requires the appropriate error variance. This is done by using the following MEANS statement after the TEST statement:

```
means A_inoc / lsd e=block*A_inoc;
```

or

```
Contrast 'Example' A_inoc 1 -1 /e=block*A_inoc;
```

SAS Output:

```
T tests (LSD) for variable: YIELD
Alpha= 0.05 df= 5 MSE= 2.31
Critical Value of T= 2.57
Least Significant Difference= 1.1271
```

T Grouping	Mean	N	A_INOC
A	20.575	24	0
B	15.950	24	1

Note that the MSE used (2.31) is the **block*A_inoc** mean square. In this particular case this test is useless because there are only two main plots. It is included only as an example.

12. 6. 4. 2. Subplot comparisons in the absence of interaction

To compare subplots it is not necessary to specify ERROR B because it is the residual. The SAS statement is:

```
means B_space / lsd;
```

SAS Output:

```
T tests (LSD) for variable: YIELD Alpha= 0.05 df= 30 MSE= 0.784
Critical Value of T= 2.04 Least Significant Difference= 0.738
```

T Grouping	Mean	N	B_SPACE
A	19.3333	12	6
A	18.8583	12	12
B	17.8833	12	4
C	16.9750	12	18

Note the different MSE used.

If the **interaction** between main plot * subplot is **significant** we are then interested in the simple effects that will result in different types of comparisons:

12.6.4.3. Among subplot treatments within common main plot

When the interaction between when A*B interaction is significant the most usual mean comparison is subplot effect within the different main plots factor. In the previous example, there is a significant interaction between main plot and subplot effects. Since the interaction is significant it is appropriate to analyze the simple effects, corresponding to point 3 above. The simple effects of most interest are those among the four spacing treatments within each inoculation treatment that can be calculated using SAS:

```
proc sort data=split;
  by A_inoc;
proc glm;
  class block B_space;
  model yield=block B_space;
  means B_space/ lsd;
  by A_inoc;
```

Output: The ANOVAS for Inoculation=0 and Inoculation =1 showed significant differences among spacings. The LSD contrasts obtained in each case are:

----- A_INOC=0 -----			
T tests (LSD) for variable: YIELD			
Alpha= 0.05 df= 15 MSE= 0.846556 (MS error for ANOVA Inoc=0)			
Critical Value of T= 2.13			
Least Significant Difference= 1.1322			
T Grouping	Mean	N	B_SPACE
A	21.5833	6	12
A	20.9833	6	18
A	20.8167	6	6
B	18.9167	6	4
----- A_INOC=1 -----			
T tests (LSD) for variable: YIELD			
Alpha= 0.05 df= 15 MSE= 0.720444 (MS error for ANOVA Inoc=1)			
Critical Value of T= 2.13			
Least Significant Difference= 1.0445			
T Grouping	Mean	N	B_SPACE
A	17.8500	6	6
B A	16.8500	6	4
B	16.1333	6	12
C	12.9667	6	18

12.6.4.4. Among main plot treatments within common subplots

Another possible comparison will be among main plots across a common subplot. The following SAS statements can be added to the previous program to test the differences between the different inoculations within each spacing treatment:

```
proc sort data=split;
  by B_space;
proc glm;
  class block A_inoc;
  model yield=block A_inoc;
  means A_inoc/ lsd;
  by B_space;
```

Note that the error here will automatically be block*A which is the correct error for main plot comparisons.

12. 6. 4. 5. Mixed comparisons: The comparison of subplot means between different main plots is more difficult because the standard error is a weighted average of MSE_A and MSE_B with emphasis on MSE_B . These comparisons require hand computations.

The appropriate MS_{AB} is:

$$MS_{AB} = \sqrt{\frac{(b-1) * MSE_B + MSE_A}{b * r}} = \sqrt{\frac{(4-1) * 0.78 + 2.31}{4 * 6}} = 0.44$$

For the mean comparisons we need an intermediate t value between the t value for the main plot ($t_{A, 5 \text{ df}} = 2.571$) and that for the subplot ($t_{B, 30 \text{ df}} = 2.042$). The formula to calculate this average t value is presented below (ST&D page 404).

$$t_{AB} = \frac{(b-1) * t_B * MSE_B + t_A * MSE_A}{(b-1)MSE_B + MSE_A} = \frac{3 * 2.042 * 0.78 + 2.571 * 2.31}{3 * 0.78 + 2.31} = 2.305$$

Note that this t_{AB} value is between t_A and t_B

Now the LSD can be calculated as

$$LSD_{\alpha=0.05} = t_{ab} * \sqrt{2} * \sqrt{\frac{(b-1) * MSE_B + MSE_A}{b * r}} = 2.305 * 1.4142 * 0.44 = 1.434$$

If the absolute value of the difference between the means being compared is larger than the critical value H_0 is rejected (there are significant differences between the subplot means in the different main plots).

For example if we want to compare the mean of inoculated / spacing 4= 16.85 with the mean of not inoculated/ spacing 6= 20.82

$|20.82-16.85|= 4.32$ and since $4.32 > \mathbf{1.434}$ (LSD) \Rightarrow This difference is significant

12.7. Split-split plot design

In testing the interaction of three factors one can arrange them in a hierarchy of subplots. The addition of a third factor by splitting subplots of a split-plot design results in a **split-split plot design**. This technique is often quite useful for a three-factor experiment to facilitate field operations or when it is desirable to keep treatment combinations together. However, the additional restriction on randomization makes it

necessary to compute a **third error term** that is used to test for main effects of the factor applied to the second split and for all interactions involving this factor. The arrangement may have certain advantages in physical operations with the experimental units, but the necessity for the third error term can make mean separation quite complicated.

The randomization procedure is the same as for the split-plot design, with the subplots being split into sub-subplots, equal in number to the levels of factor three, to which the third factor is randomly assigned- a new randomization for each set of sub-subplots.

The following figure from Little & Hills, illustrates the partial layout of a split-split plot to evaluate the effects of dates of planting (A), aphid control (B), and date of harvest (C) on the control of aphid-borne sugar beet viruses.

Block I			A ₁			II A ₃			A ₂		
A ₃ B ₁	III A ₁ B ₁	A ₂ B ₂	A ₂ B ₁ C ₁	A ₂ B ₁ C ₃	A ₂ B ₁ C ₂	A ₁ B ₂ C ₃	IV A ₁ B ₂ C ₁	A ₁ B ₂ C ₂	A ₃ B ₂ C ₁	A ₃ B ₂ C ₃	A ₃ B ₂ C ₂
A ₃ B ₂	A ₁ B ₂	A ₂ B ₁	A ₂ B ₂ C ₃	A ₂ B ₂ C ₂	A ₂ B ₂ C ₁	A ₁ B ₁ C ₁	A ₁ B ₁ C ₃	A ₁ B ₁ C ₂	A ₃ B ₁ C ₃	A ₃ B ₁ C ₁	A ₃ B ₁ C ₂

The analysis of variance is basically a straightforward extension of the split-plot. The different error terms can be constructed by pooling together different sources of variation.

Block
A
Block * A = Error A
B
A*B
Block * B + Block * A * B = Block*B(A) = Error B
C
A*C
B*C
A*B*C
Block * C + Block * A * C + Block * B * C + Block * A * B * C = residual = Error C

The following SAS statements produce an analysis of variance and correct tests.

```
proc glm;
  classes Block a b c;
  model response= Block a Block*a b a*b Block*b*a c a*c b*c a*b*c;
  test h=a e=Block*a;
  test h=b e=Block*b*a;
  test h=a*b e=Block*b*a;
```

The Sum of Squares **Block*b*a** is computationally equivalent to the Sum of Squares **Block*b + Block*a*b**. A feature of PROC GLM is that MODEL statements containing interaction terms without one or more of the corresponding lower-effect terms listed in the model, produce Sum of Squares that contain the nonspecified main-effect sums of squares. For example the SAS statements:

model y= a*b → produces a SS labeled a*b that is actually SS(A)+SS(B)+SS(A*B)

model y= a a*b → produces a SS labeled a*b that is actually SS(B)+SS(A*B)

model y= Block*a*b → produces a SS labeled Block*a*b that is actually SS(Block*b) + SS(Block*a*b), because all the factors in the Block*b interaction are within Block*a*b and the Block*b interaction is not listed in the model statement.

A complete analysis of a split-split plot example is discussed by Little and Hills (1978 Chapter 9).

12. 8. Split-block (or strip-plot) design

In the strip-plot or split-block design the subunit treatments are applied in strips across an entire replication of main plot treatments.

Here is a comparison of the layout for a 5 x 4 split-plot and a 5 x 4 split-block (only one replication is shown). Although the terms main plot and subplot are still used, from a theoretical perspective there is no difference between the two (i.e., they are symmetric).

A3	A2	A1	A5	A4
B2	B1	B2	B3	B4
B1	B3	B1	B2	B3
B3	B2	B4	B4	B1
B4	B4	B3	B1	B2
Split-plot				

A3	A2	A1	A5	A4
B2	B2	B2	B2	B2
B4	B4	B4	B4	B4
B1	B1	B1	B1	B1
B3	B3	B3	B3	B3
Split-block or Strip-plot				

Note that the subunit treatments are continuous across the entire block or main plot, and thus each subunit treatment splits the block. Another term applicable to this layout is **strip-plot**, as both A and B treatments are in strips. The A and B treatments are independently randomized in each replication.

12. 8. 1. Reasons for doing a split-block design

- 1) Physical operations (e.g. tractor manipulation, irrigation, harvesting) may be easier.
- 2) The design tends to sacrifice precision in the main effects and **improve precision in the interaction effects**. This may be the most important part of the experiment.

12. 8. 2 Linear model for the split-block design

The linear model for the split-block with RCB main plots (this is the most common design) is

$$Y_{ijk} = \mu + \rho_k + \alpha_i + \beta_j + \gamma_{ik} + \theta_{jk} + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

where $k = 1, \dots, r$ indexes the blocks, $i = 1, \dots, a$ indexes the main plot levels, and $j = 1, \dots, b$ indexes the subplot levels. The extra term θ_{jk} represents the interaction of blocks with subplot levels.

This θ_{jk} term was not specified in the split-plot model and consequently its variation was included in the Subplot error, MS (SPE), estimated from the interaction [*Subplot x Replication + Main plot x Subplot x Replication*] (see topic 12.5).

12. 8. 3. ANOVA for the split-block design

The strip-plot error (MS (STPE)= *Subplot x Replication*) is the denominator of the F statistic for the main effects of factor B. This test is symmetric to the test for factor A, where $SS(MPE)=$ *Main plot x Replication* is the denominator of the F test. This is a reasonable result considering that in the split-block design both factors have similar randomization and are symmetric. Another way to think about this error term is to consider the average of all main plots in each subplot. This will result in a RCBD for factor B with one replication per cell. The appropriate error term in this case is the MS of *Factor B x Block*.

The ANOVA table for the RCBD split block design is

Source	df	SS	MS	F
Blocks	$r - 1$	SS(Block)		
Factor A	$a - 1$	SSA	MSA	MSA / MS(MPE)
Error A= A*block	$(a - 1)(r - 1)$	SS(MPE)	MS(MPE)	
Factor B	$b - 1$	SSB	MSB	MSB / MS(STPE)
Error B= B*block	$(b - 1)(r - 1)$	SS(STPE)	MS(STPE)	
A x B	$(a - 1)(b - 1)$	SS(AxB)	MS(AxB)	MS(AxB) / MS(SPE)
Error C=A*B*block	$(a-1)(r-1)(b-1)$	SS(SPE)	MS(SPE)	
Total (subplots)	$rab - 1$	SS		

This new error term, MS (STPE), is subtracted from the subplot error (Error C) and it takes $(r-1)(b-1)$ degrees of freedom from this error. The result of this subtraction is a smaller error C, that is the error term used to test the interaction AxB. This results in an **improved precision in the tests for interaction effects.**

12. 8. 4. Example of split-block (modified from Little and Hills, Chapter 10)

The following figure gives the layout of an experiment designed to examine the effect of nitrogen fertilizer rate on sugar beet root yield for different harvest times. The main plots are four nitrogen fertilizer rates arranged in a RCBD with two blocks. Subunit treatments are five dates of harvest. The subplots to be harvested at each date are in strips through an entire column of main plots. The harvest date strips are randomized for each column of main plots. Harvest operations are easier to conduct when the plots to be harvested on a certain date form a continuous column.

The root yield in tons per acre for each subplot, and the means for each N main plot are given in the next figure:

Block I					Block II				
H4	H5	H1	H3	H2	H4	H2	H3	H5	H1
N 80 (plot total = 107.1)					N 160 (plot total = 116.2)				
26.4	29.3	10.1	23.1	18.2	34.2	18.5	22.4	30.3	10.8
N 320 (plot total = 120.8)					N 0 (plot total = 74.8)				
31.2	34.2	10.3	25.9	19.2	21.3	12.5	16.7	19.1	5.2
N 160 (plot total = 109.6)					N 80 (plot total = 102.9)				
28.0	31.2	10.2	22.3	16.9	29.5	16.9	20.4	26.6	9.5
N 0 (plot total = 42.4)					N 320 (plot total = 109.1)				
10.1	11.4	2.3	9.8	8.8	31.9	17.8	22.8	29.2	7.4

12. 8. 4. 1. SAS Program for a split block

```

data splitblk;
input block H_B N_A yield @@;
cards;
1 1 0 2.3 1 2 0 8.8 1 3 0 9.8 1 4 0 10.1 1 5 0 11.4
1 1 80 10.1 1 2 80 18.2 1 3 80 23.1 1 4 80 26.4 1 5 80 29.3
1 1 160 10.2 1 2 160 16.9 1 3 160 22.3 1 4 160 28.0 1 5 160 31.2
1 1 320 10.3 1 2 320 19.2 1 3 320 25.9 1 4 320 31.2 1 5 320 34.2

2 1 0 5.2 2 2 0 12.5 2 3 0 16.7 2 4 0 21.3 2 5 0 19.1
2 1 80 9.5 2 2 80 16.9 2 3 80 20.4 2 4 80 29.5 2 5 80 26.6
2 1 160 10.8 2 2 160 18.5 2 3 160 22.4 2 4 160 34.2 2 5 160 30.3
2 1 320 7.4 2 2 320 17.8 2 3 320 22.8 2 4 320 31.9 2 5 320 29.2
;
proc GLM;

```

```

class block N_A H_B;
model yield = block
            N_A N_A*block
            H_B H_B*block H_B*N_A;
test h=N_A e=N_A*block;
test h=H_B e=H_B*block;
run;

```

Output

```

BLOCK          2      1 2
N_A            4      0 80 160 320
H_B           5      1 2 3 4 5
Number of observations in data set = 40

```

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	3027.2658	112.1210	88.67	0.0001
Error	12	15.1740	1.2645		
Corrected Total	39	3042.4398			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	1	14.5203	14.5203	11.48	0.0054
N_A	3	838.2988	279.4329	220.98	0.0001
BLOCK*N_A	3	111.6847	37.2282	29.44	0.0001
H_B	4	1898.9460	474.7365	375.43	0.0001
BLOCK*H_B	4	42.7860	10.6965	8.46	0.0017
N_A*H_B	12	121.0300	10.0858	7.98	0.0005

Tests of Hypotheses using the Type III MS for BLOCK*N_A as an error term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
N_A	3	838.29875	279.43292	7.51	0.0660

Tests of Hypotheses using the Type III MS for BLOCK*H_B as an error term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
H_B	4	1898.9460	474.7365	44.38	0.0014

The F tests from the main ANOVA table obtained using inappropriate MSE were strikethrough. The correct tests are provided below as result of the TEST statements. There are no significant differences among Nitrogen levels but differences among harvest dates are highly significant. The interaction Nitrogen * Harvest date is also highly significant.

The F test for the nitrogen levels is almost significant. Since four level of nitrogen were tested, the resulting SS includes a linear, a quadratic and a cubic response. It is not easy to write a contrast for these effects because the levels of nitrogen selected are not equally spaced. However, as we exemplified in Laboratory 4&5, this problem can be overcome using a multiple regression approach. The following simplified program can be used to partition the Nitrogen sum of squares:

```
proc GLM;
```

```
class block;
model yield = block N_A N_A*N_A N_A*N_A*N_A;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	1	14.52025	14.52025	0.23	0.6330
N_A	1	508.20875	508.20875	8.12	0.0073
N_A*N_A	1	290.18977	290.18977	4.64	0.0382
N_A*N_A*N_A	1	39.90023	39.90023	0.64	0.4299

The addition of the linear effect (508.21), the quadratic effect (290.19), and the cubic effect (39.90) equals the total sum of squares for Nitrogen (838.30) from the previous model.

The appropriate F tests show a significant ($p < 0.05$) lineal effect.

Source	DF	Type I SS	Mean Square	F Value
N_A	1	508.20875	508.20875	13.65 *
N_A*N_A	1	290.18977	290.18977	7.79 NS
N_A*N_A*N_A	1	39.90023	39.90023	1.07 NS
BLOCK*N_A	3	111.6847	37.2282	

The study of simple effects of Nitrogen at each Harvest date or the simple effects of Harvest date at each Nitrogen level would be an appropriate continuation of this study.

The random statement shows the correct strip-plot errors

```
Model Yield = Block A Block*A B Block*B A*B;
Random Block Block*A Block*B;
```

Source	Type III Expected Mean Square
Block	Var(Error) + 2 Var(Block*B) + 3 Var(Block*A) + 6 Var(Block)
A	Var(Error) + 3 Var(Block*A) + Q(A,A*B)
Block*A	Var(Error) + 3 Var(Block*A)
B	Var(Error) + 2 Var(Block*B) + Q(B,A*B)
Block*B	Var(Error) + 2 Var(Block*B)
A*B	Var(Error) + Q(A*B)

Block*A is the correct error for A

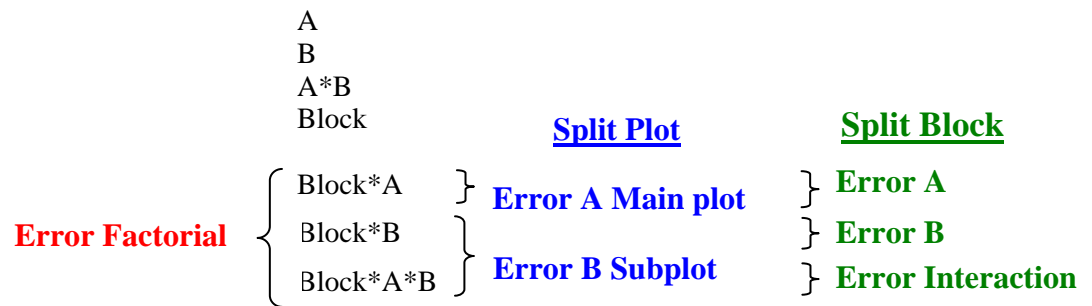
Block*B is the correct error for B

A synthetic error is necessary to test blocks, if necessary.

Review: Pooling of interaction components to construct Split-plot, Split-split-plots and Split Block error terms

Two way factorial

Class Block A B;
Model Y= Block A B A*B;



Three way factorial

Class Block A B;
Model Y= Block A B A*B

